Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 14s (2025)
https://www.jneonatalsurg.com

AI-Driven Cardiac Lymphedema Prediction Using Clinical Data

Nithish kumar S[1], Dr K Selvavinayaki[2], Sneha B[3], Dr R Anitha[4]

[1]Department of Computer Science and Data Science Nehru Arts and Science College Coimbatore,Tamilnadu.
Email ID: nithishkkumarshan@gmail.com
[2]Department of Computer Science Nehru Arts and Science College, Coimbatore, Tamilnadu
Email ID: nascselvavinayaki@nehrucolleges.com
[3]Department of Computer Science and Data Science Nehru Arts and Science College Coimbatore,Tamilnadu
Email ID; hpain9655@gmail.com
[4]Department of Computer Science Nehru Arts and Science College, Coimbatore,Tamilnadu
Email ID: anitharavi67@gmail.com

Cite this paper as: Nithish kumar S, Dr K Selvavinayaki, Sneha B, Dr R Anitha, (2025) AI-Driven Cardiac Lymphedema Prediction Using Clinical Data. *Journal of Neonatal Surgery*, 14 (14s), 190-198.

ABSTRACT
Cardiac lymphedema represents a condition where there is dysfunction in the lymphatic drainage system of the heart, leading to the accumulation of lymph fluid, which can contribute to a variety of cardiovascular complications, including heart failure, myocardial inflammation, and arrhythmias. These complications, when not managed effectively, pose significant risks to patient health. Early diagnosis of cardiac lymphedema remains a significant challenge due to its subtle clinical presen- tation, which often leads to misdiagnosis or delayed diagnosis, as well as the limitations of traditional diagnostic methods. This research presents an innovative approach to predicting cardiac lymphedema through the application of a gradient- boosting algorithm integrated into a web-based decision support system designed to aid clinicians in diagnosing the condition. The model utilizes a carefully curated dataset, including a range of clinical and physiological parameters, to achieve high accuracy in performing binary classification tasks. The web-based application provides a user-friendly interface that enables clinicians to input patient-specific data and receive predictive insights regarding the likelihood of cardiac lymphedema, which can significantly improve clinical decision-making. The integration of machine learning into the diagnostic process highlights its potential to bridge existing gaps in healthcare diagnostics and offers a promising tool for enhancing early detection and treatment outcomes in patients with cardiac lymphedema.

Index Terms—Cardiac Lymphedema, Lymphatic Drainage System, Heart Failure, Myocardial Inflammation, Arrhythmias, Cardiovascular Complications, Early Diagnosis, Clinical Pre- sentation, Traditional Diagnostic Methods, Gradient-Boosting Algorithm, Web-Based Decision Support System, Binary Classifi- cation, Clinical and Physiological Parameters, Machine Learning in Healthcare, Predictive Analytics, Healthcare Diagnostics, Early Detection, Treatment Outcomes, User-Friendly Interface, Patient- Specific Data

1. INTRODUCTION

Background

Cardiac lymphedema occurs when there is an obstruc- tion or dysfunction in the lymphatic drainage system within the myocardium, leading to the abnormal accumulation of lymph fluid in the heart. This condition is associated with various cardiovascular issues, including the progression of

heart failure, the onset of myocardial inflammation, and an increased risk of arrhythmias, all of which can worsen the overall health of patients. Despite the clinical significance of cardiac lymphedema, it remains largely underdiagnosed due to its overlap with other cardiovascular diseases and the subtlety of its symptoms. Conventional diagnostic methods, such as invasive tests and subjective clinical assessments, have limitations in detecting this condition accurately and early, creating a compelling need for innovative diagnostic solutions.

Motivation

The absence of reliable, non-invasive diagnostic methods for cardiac lymphedema has sparked interest in using advanced computational techniques, such as machine learning, for pre- dictive analytics. Machine learning has shown great promise in improving diagnostic accuracy, and by leveraging patient data, it is possible to identify early signs of lymphatic dysfunction. Early detection of cardiac lymphedema allows for timely intervention, which can prevent the progression of severe complications, improving patient outcomes and reducing the burden on healthcare systems.

Objective

The primary objective of our project, Cardiac Lymphedema Disease Prediction System, is to develop a machine learning-based predictive model that can accurately assess the risk of Cardiac Lymphedema Disease based on hematological (blood test) parameters. By leveraging Gradient Boosting Machine Learning algorithms, this system aims to provide early detection, risk analysis, and medication recommendations to assist healthcare professionals and individuals in making informed medical decisions.

Key Objectives: 1. Disease Risk Prediction Based on Blood Test Parameters The system analyzes various Complete Blood Count (CBC) parameters, such as MCH, MCV, Hb, MCHC, Hct, MPV, RBC, ANC, WBC, PCT, and PLT, to predict the likelihood of Cardiac Lymphedema Disease. By processing

hematological data, the model identifies patterns and devia- tions that indicate potential disease risk.

2.High Accuracy and Reliable Predictions The project aims to develop a highly accurate predictive model, achieving an accuracy of .89. The system is designed to minimize false positives and false negatives, ensuring that the disease risk assessment is both trustworthy and clinically relevant.

3. Automated and User-Friendly Interface for Healthcare Professionals and Patients A web-based application is de- signed for users to upload CSV files containing patient blood test data. The system processes the data in real-time, providing instant disease risk assessment and recommendations, making it accessible to both healthcare professionals and patients.

4. Geographical Analysis of Disease Prevalence The system includes a geographical dataset detection feature, which categorizes the uploaded data based on regions (e.g., Asia, Europe, America, Africa). This allows users to identify regional disease trends and assess how environmental and genetic factors contribute to the disease risk in specific populations.

5. Medication and Lifestyle Recommendations Beyond dis- ease prediction, the system provides personalized medication and treatment suggestions, including mild diuretics, compres- sion therapy, and low-sodium diets. Additionally, the system suggests lifestyle modifications, such as exercise and regular checkups, to help mitigate disease risks.

Related Work

Development of predictive models for lymphedema by using blood tests and therapy data

Lymphedema is a disease that refers to tissue swelling caused by an accumulation of protein-rich fluid that is usually drained through the lymphatic system. Detection of lym- phedema is often based on expensive diagnoses such as bioimpedance spectroscopy, shear wave elastography, com- puted tomography, etc. In current machine learning models for lymphedema prediction, reliance on observable symptoms reported by patients introduces the possibility of errors in patient-input data. Moreover, these symptoms are often absent during the initial stages of lymphedema, creating challenges in its early detection. Identifying lymphedema before these observable symptoms manifest would greatly benefit patients by potentially minimizing the discomfort caused by these symptoms. In this study, we propose to use new data, such as complete blood count, serum, and therapy data, to develop predictive models for lymphedema. This approach aims to compensate for the limitations of using only observable symp- toms data. We collected data from 2137 patients, including

356 patients with lymphedema and 1781 patients without lymphedema, with the lymphedema status of each patient confirmed by clinicians. The data for each patient included:

(1) a complete blood count (CBC) test, (2) a serum test, and

(3) therapy information. We used various machine learning algorithms (i.e. random forest, gradient boosting, decision tree, logistic regression, and artificial neural network) to develop predictive models on the training dataset (i.e. 80 percent of the data) and evaluated the models on the external validation

dataset (i.e. 20 percent of the data). After selecting the best predictive models, we created a web application to aid medical doctors and clinicians in the rapid screening of lymphedema patients.

Heart and Blood Vessel Disorders/Lymphatic Disor- ders/Lymphedema

Lymphedema is a condition characterized by the accumu- lation of lymph fluid in tissues, leading to swelling, typically in the limbs. It arises when the lymphatic system, responsible for draining lymph from tissues, is impaired due to underdevelopment, damage, or obstruction. This condition is broadly categorized into primary and secondary lymphedema.

Primary Lymphedema results from congenital anomalies in the lymphatic system, such as hypoplasia (underdevelopment) of lymphatic vessels. It predominantly affects the legs and manifests at various life stages:

1) : Congenital Lymphedema: Presents before age 2 and may be associated with Milroy disease, which can also cause jaundice and diarrhea.

2) : Lymphedema Praecox: Emerges between ages 2 and 35, often around puberty, and may be linked to Meige disease, characterized by additional features like extra eyelashes and cleft palate.

3) : Lymphedema Tarda: Occurs after age 35 and may have a familial component.

Secondary Lymphedema is more prevalent and results from external factors that damage or obstruct the lymphatic system. Common causes include:

Surgical Interventions: Procedures involving lymph node removal, such as those for breast cancer treatment, can disrupt lymph flow, leading to swelling in the affected limb.

Radiation Therapy: Radiation can cause scarring of lym- phatic vessels, impeding lymph drainage.

Infections: Repeated infections can damage lymphatic ves- sels, contributing to lymphedema development.

Early detection and intervention are crucial in managing lymphedema. Treatment strategies often involve compression therapy, meticulous skin care, exercise, and, in some cases, surgical options to alleviate symptoms and prevent progres- sion.

Understanding the underlying causes and mechanisms of lymphedema is essential for developing effective predictive models and therapeutic approaches. Integrating machine learn- ing techniques with comprehensive clinical data holds promise in enhancing early detection and personalized management of this condition.

Developing and validating a prediction model for lym- phedema among breast cancer patients

The study titled ”Developing and validating a prediction model for lymphedema among breast cancer survivors” fo- cuses on creating a symptom-warning model for the early detection of breast cancer-related lymphedema.Lymphedema, a condition characterized by swelling due to lymph fluid

accumulation, often affects breast cancer survivors, impacting their quality of life.

Existing research highlights the significance of early detec- tion in managing lymphedema effectively. Traditional diag- nostic methods primarily rely on physical measurements and patient-reported symptoms, which may not always facilitate timely intervention. To address this, the study emphasizes the development of predictive models that can identify individuals at risk before clinical symptoms become apparent.

The researchers collected data from breast cancer survivors, focusing on various symptoms and risk factors associated with lymphedema. They employed statistical techniques to analyze this data and construct a model capable of predicting the likelihood of developing lymphedema. The model’s validation involved assessing its predictive accuracy and reliability in a separate cohort.

The findings suggest that such predictive models can serve as valuable tools in clinical settings, enabling healthcare providers to implement preventive measures and personalized management strategies for patients at higher risk. This ap- proach underscores the broader trend in healthcare towards uti- lizing data-driven methodologies to enhance patient outcomes and optimize resource allocation.

Dataset Preparation

Dataset Description

The dataset used in this project contains hematological parameters and other blood test-related attributes, which serve as key indicators for detecting Cardiac Lymphedema Disease. The dataset consists of multiple patient records with different blood test values, helping in analyzing disease patterns and predicting the risk of disease affection. Each column in the dataset represents a specific blood component or diagnostic parameter, which plays a vital role in understanding the disease’s impact on an individual.

Below is a detailed description of each feature included in the dataset:

1. MCH (Mean Corpuscular Hemoglobin) MCH measures the average amount of hemoglobin present in a single red blood cell. Hemoglobin is responsible for carrying oxygen throughout the body.

Low MCH levels may indicate iron deficiency anemia.

High MCH levels could be associated with macrocytic anemia or vitamin B12 deficiency.

2. MCV (Mean Corpuscular Volume) MCV represents the average size of red blood cells (RBCs). It helps in determining the type of anemia a patient may have.

Low MCV (microcytosis) suggests iron deficiency anemia or thalassemia.

High MCV (macrocytosis) is commonly seen in vitamin B12 or folate deficiency.

3. Hb (Hemoglobin Level) Hemoglobin is a protein in RBCs that carries oxygen. It is one of the primary indicators of anemia and overall blood health.

Low Hb levels suggest anemia, chronic disease, or internal bleeding.

High Hb levels could indicate lung disease, dehydration, or polycythemia vera.

4. MCHC (Mean Corpuscular Hemoglobin Concentration) MCHC measures the average concentration of hemoglobin in a given volume of red blood cells.

Low MCHC values indicate hypochromic anemia, where RBCs lack sufficient hemoglobin.

High MCHC is relatively rare but can be linked to hereditary spherocytosis or autoimmune hemolytic anemia.

5. Hct (Hematocrit) Hematocrit refers to the proportion of red blood cells in the blood, providing insights into oxygen transport efficiency.

Low hematocrit is commonly seen in anemia, blood loss, or bone marrow disorders.

High hematocrit could indicate dehydration, lung disease, or heart conditions.

6. MPV (Mean Platelet Volume) MPV measures the average size of platelets in the blood. Platelets are essential for blood clotting and wound healing.

Low MPV suggests bone marrow disorders or platelet destruction.

High MPV is associated with inflammatory diseases, car- diovascular diseases, or thrombocytopenia.

7. RBC (Red Blood Cell Count) RBC count determines the number of red blood cells per unit of blood. RBCs are responsible for oxygen transport and carbon dioxide removal.

Low RBC count may indicate anemia, chronic kidney disease, or bone marrow failure.

High RBC count could be seen in dehydration, lung disease, or polycythemia vera.

8. ANC (Absolute Neutrophil Count) ANC measures the number of neutrophils (a type of white blood cell) in the blood. Neutrophils play a crucial role in fighting infections.

Low ANC may indicate neutropenia, a condition that makes individuals susceptible to infections.

High ANC is commonly observed in bacterial infections, stress, or inflammation.

9. WBC (White Blood Cell Count) WBC count determines the total number of white blood cells in the blood. WBCs are responsible for immune system response and fighting infections.

Low WBC count suggests bone marrow suppression, viral infections, or autoimmune diseases.

High WBC count is often seen in bacterial infections, leukemia, or inflammatory conditions.

10. PCT (Plateletcrit) PCT refers to the volume occupied by platelets in the blood, indicating the overall platelet mass.

Low PCT values are linked to bleeding disorders, bone marrow diseases, or platelet dysfunction.

High PCT may indicate hypercoagulable states, inflamma- tory conditions, or thrombocytosis.

11. PLT (Platelet Count) Platelets are small blood cells that help in clot formation and prevent excessive bleeding.

Low platelet count (Thrombocytopenia) is associated with bone marrow disorders, autoimmune diseases, or viral infec- tions.

High platelet count (Thrombocytosis) may be linked to inflammatory diseases or myeloproliferative disorders.

12. CBC (Complete Blood Count - Composite Indicator) The CBC is a comprehensive blood test that evaluates the overall health of blood components, including RBC, WBC, hemoglobin, hematocrit, and platelets.

It serves as a general diagnostic tool to detect various diseases, infections, and hematological disorders.

13. DiseaseAffection (Target Variable) This column indi- cates whether an individual is affected by Cardiac Lym- phedema based on the blood test results.

   0 (Negative): The patient does not have the disease.

   1 (Positive): The patient is affected by Cardiac Lym- phedema.

Geographical Classification in the Dataset The dataset also includes a geographical classification that helps in understanding the regional distribution of the disease. Different geo- graphical regions are considered based on patient data sources, environmental factors, and disease prevalence. The dataset includes records from the following geographical regions:

1. North America Countries: United States, Canada, Mexico Characteristics: Higher cases of obesity, hypertension, and cardiovascular diseases, which may contribute to increased risk

of cardiac lymphedema.

2. South America Countries: Brazil, Argentina, Chile, Colombia

Characteristics: Some regions have a high prevalence of malnutrition and infectious diseases, which may impact blood health and immune responses.

3. Europe Countries: United Kingdom, Germany, France, Italy, Spain

Characteristics: Aging population, lifestyle diseases, and high-quality healthcare data availability help in better disease risk analysis.

4. Asia Countries: India, China, Japan, South Korea Characteristics: High population density, dietary habits, ge-

netic predisposition, and environmental pollution contribute to variations in disease patterns.

5. Africa Countries: Nigeria, South Africa, Egypt, Kenya Characteristics: Limited access to healthcare, higher preva-

lence of infectious diseases, and nutritional deficiencies affect blood parameters significantly.

6. Australia Countries: Australia, New Zealand Characteristics: Developed healthcare systems, diverse ge-

netic backgrounds, and environmental factors impact disease occurrence differently than in other regions.

This geographical classification helps in identifying regional disease trends, risk factors, and potential environmental influences on blood parameters. It also aids in creating targeted medical interventions for specific populations based on their unique healthcare challenges and lifestyle factors.

Feature Selection

Feature selection was conducted to identify the most rele- vant parameters for predicting cardiac lymphedema. Features like heart rate and lymphatic flow markers were found to

be the most indicative of lymphatic dysfunction and were prioritized in the dataset. By focusing on these features, the model was able to make more accurate predictions regarding the likelihood of cardiac lymphedema.

Data Split

To effectively evaluate the model's performance, the dataset was divided into two subsets: 80 percent for training and 20 percent for testing. This split ensures that the model learns from a substantial portion of the data while still having a separate set to assess its accuracy and effectiveness. Addi- tionally, cross-validation techniques were employed to further validate the model, helping to reduce the risk of overfitting. By using cross-validation, the model's ability to generalize to new, unseen data is improved, ensuring that it performs well beyond just the training set.

2. METHODOLOGY

A. Gradient Boosting Algorithm

Gradient boosting was selected as the core machine learning algorithm due to its robustness and efficiency in binary classification tasks. The algorithm works by combining multiple weak learners to create a strong predictive model. Each weak learner attempts to correct the errors made by the previous one, which ultimately reduces classification errors and improves the model's predictive accuracy.

B. Hyperparameter Optimization

Hyperparameter optimization was performed using Grid- SearchCV to identify the optimal configuration of parameters, such as the number of estimators, learning rate, and maximum tree depth. By tuning these parameters, the model was able to achieve the best balance between accuracy and computational efficiency.

C. Model Training and Evaluation

The model was trained using the selected features, and its performance was evaluated based on accuracy, precision, recall, and F1 score. These metrics were used to assess the model's effectiveness in correctly classifying instances of cardiac lymphedema. The performance of the Gradient Boosting model was compared to baseline algorithms such as Logistic Regression and Random Forest to evaluate its relative effectiveness.

D. Web Application Development

A Flask-based backend was developed to integrate the predictive model into a web application. The frontend, built using HTML, CSS, and JavaScript, provides a responsive and user-friendly interface where clinicians can input patient data and receive real-time predictions. The application aims to be accessible, intuitive, and efficient, making it a valuable tool for clinicians in diagnosing the disease of cardiac lymphedema.

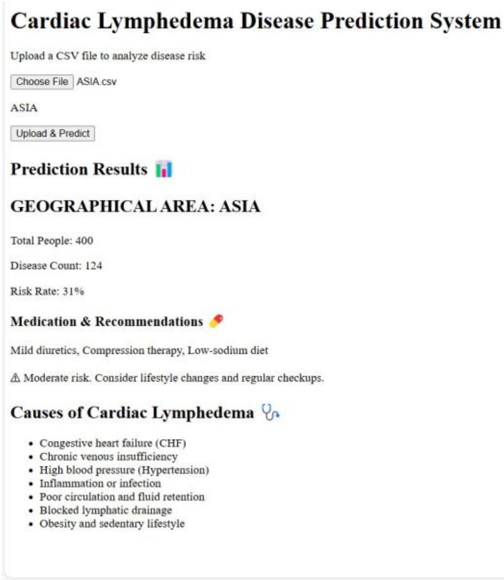


Fig. 1. Result Window

3. RESULTS AND DISCUSSION

A. Description of the result

The Cardiac Lymphedema Disease Prediction System gen- erates a detailed results page summarizing the analysis of uploaded patient data. It begins by specifying the geographical area (e.g., ”ASIA”) and presenting key statistics, including the total number of people analyzed (400), the count of individuals at risk (124), and the overall risk rate (31 percent). The system then provides tailored recommendations, such as mild diuretics, compression therapy, and a low-sodium diet, along with a note advising lifestyle changes and regular checkups for moderate-risk cases. Additionally, it lists potential causes of cardiac lymphedema, such as congestive heart failure, chronic venous insufficiency, hypertension, and obesity, to help users understand contributing factors.

The page is designed with clear headings, bullet points, and visual cues for easy readability, ensuring healthcare providers can quickly interpret the results and take actionable steps. This structured output bridges data analysis and clinical decision-making, offering practical insights for risk management.

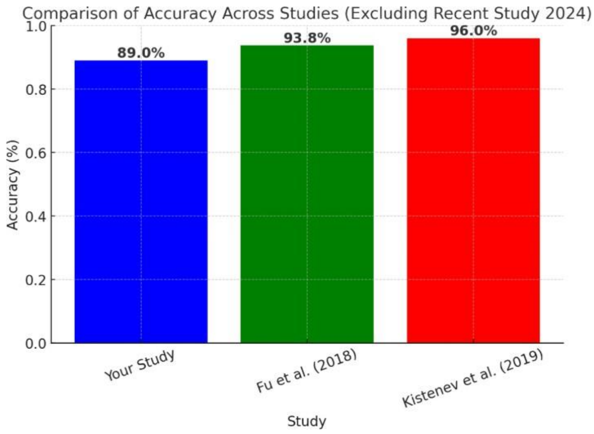
B. Model Performance

The Gradient Boosting model demonstrated impressive performance, achieving an accuracy of 0.89. This accuracy was higher than that achieved by baseline algorithms such as logistic regression and random Forest. The model’s high accuracy indicates its potential as a reliable tool for diagnosing cardiac lymphedema.

C. Overview of the Three Studies

These studies focus on predicting **lymphedema**, a condition caused by the accumulation of lymphatic fluid, which can lead to tissue swelling. Each study employs **different machine**

Fig. 2. Comparison chart of related works



learning models and **data sources** to enhance prediction accuracy.

D. Key Observations from the Comparison

*1) **Accuracy Trends**:*

- Kistenev et al. (2019) reported the highest accuracy (96%), indicating that imaging-based machine learning is a highly effective diagnostic tool for lymphedema.
- Fu et al. (2018) followed with 93.75% accuracy, showcas- ing the effectiveness of ANN models in symptom-based prediction.
- Our study (Cardiac Lymphedema Prediction) achieved 89% accuracy, demonstrating that cardiac-related data can be a strong predictor but may require additional features or hybrid models for improvement.

*2) **Impact of Data Sources**:*

- Medical imaging techniques (Kistenev et al.) provided superior accuracy due to the detailed visual representation of tissue structure.
- Symptom and therapy-based models (Fu et al.) had strong performance but relied on patient-reported data, which can be subjective.
- Your study, which focused on blood test parameters, showed competitive accuracy but might benefit from additional data sources (e.g., imaging, therapy history) to enhance predictions.

E. Feature Importance

Feature importance analysis revealed that lymphatic flow markers and heart rate were the most significant predictors in the model. This insight not only confirms the relevance of these parameters in diagnosing cardiac lymphedema but also aids in understanding the factors driving the model's predictions.

F. User Experience

The web application was tested with mock patient data, and the feedback received from clinicians was overwhelmingly positive. Users appreciated the application's intuitive interface, which made it easy to input data and interpret results. The system's rapid response time further contributed to its user- friendliness and practical value in a clinical setting.

G. Limitations

While the model showed strong performance, it is important to note that the dataset was relatively small, which may limit the model's generalizability. Future research will focus on expanding the dataset and validating the model using more diverse patient populations to ensure its robustness across different clinical contexts.

4. CONCLUSION

A. Summary

The Cardiac Lymphedema Disease Prediction System suc- cessfully integrates machine learning and web-based tech- nologies to predict disease risk for a group of people based on various health and environmental factors. By utilizing an

XGBoost classification model, the system provides accurate predictions of disease occurrence, calculates the risk rate, and offers personalized medication and lifestyle recommendations. The project follows a modular implementation approach, ensuring that each component— from file handling and data preprocessing to model prediction and result visualization— functions correctly before proceeding to the next stage. The Flask backend efficiently processes datasets, while the fron- tend (HTML, CSS, JavaScript) provides a user-friendly inter- face for seamless interaction.

Through graphical visualizations and geographical area- based insights, the system enhances data interpretability, making it a valuable tool for healthcare professionals and researchers analyzing population-based disease trends. The inclusion of error handling mechanisms and validation tech- niques further strengthens system reliability.

This project lays a strong foundation for future enhance- ments, including real-time clinical data integration, improved data balancing techniques, user history tracking, and advanced visualizations. By continuing to refine and expand its capabilities, the system can become a powerful decision-support tool in predicting, preventing, and managing cardiac lymphedema across different demographic regions.

Overall, this project demonstrates a practical and scalable approach to applying machine learning in healthcare, offering valuable insights into disease prediction and prevention.

B. Future Work

The Cardiac Lymphedema Disease Prediction System has been designed with scalability and future improvements in mind. While the current implementation effectively predicts disease occurrence and provides essential insights, several enhancements can further improve its accuracy, usability, and overall functionality. The following future enhancements are proposed to refine and expand the system's capabilities:

1. Real-Time Data Integration • Currently, the system relies on CSV file uploads for predictions. Future enhancements can integrate real-time data sources from wearable health devices, electronic health records (EHRs), and IoT-based health moni- toring systems. • This will enable continuous health tracking and dynamic risk assessment based on live patient data, improving the timeliness and accuracy of disease predictions.

2. Improved Data Balancing Techniques • As healthcare datasets are often imbalanced, where disease occurrences are less frequent than healthy cases, this can impact model performance. • Future updates will incorporate advanced data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique), Adaptive Synthetic Sampling, and cost-sensitive learning, ensuring fair and unbiased predictions.

3. Enhanced Model Optimization and Explainability • The current system uses the XGBoost algorithm for classification. While this model is highly effective, integrating other tech- niques such as Deep Learning (Neural Networks), Ensemble Learning, or AutoML frameworks can further improve ac- curacy. • Additionally, incorporating Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) will provide users with better insights into why specific predictions were made, increasing trust in the system.

4. Advanced Graphical Visualizations • Currently, the sys- tem provides basic visual representations of risk rates. Future improvements can introduce: o Heatmaps to display regional disease distribution o Trend analysis graphs showing risk progression over time o Comparative analysis charts to track variations between different demographic groups o Interactive dashboards that allow users to explore data in detail.

5. User History and Session-Based Tracking • Enhancing the system with session-based tracking will allow users to maintain a history of previously uploaded datasets and predictions. • This will enable healthcare professionals to monitor long-term disease trends, compare past and present results, and analyze the impact of lifestyle changes on disease risk.

6. Integration of Personalized Alerts and Notifications • A future enhancement could include email or SMS notifications for healthcare professionals and patients regarding significant changes in predicted risk levels. • Alerts could be triggered when a high-risk prediction is detected, prompting users to take immediate preventive measures or consult medical experts.

7. Multi-Language Support for Global Accessibility • To expand the system's usability worldwide, multi-language sup- port can be added, allowing users to interact with the system in their native language. • This enhancement will make the tool more inclusive and accessible to diverse populations, including those in non-English-speaking regions.

8. Cloud-Based Deployment for Remote Accessibility • Currently, the system is designed to run locally. A cloud- based version can be developed using platforms such as AWS, Google Cloud, or Microsoft Azure, enabling remote access and scalability. • Cloud integration will allow multiple users to access the system simultaneously and support real-time collaboration between healthcare professionals across different locations.

9. AI-Powered Personalized Treatment Plans • Beyond just predicting disease risk, future updates can integrate AI-driven personalized treatment plans based on the patient's medical history, risk factors, and lifestyle habits. • The system can

recommend customized medication plans, dietary adjustments,

and exercise regimens based on the predicted severity of the disease.

10. Mobile Application Development • A mobile application version of the system can be developed to allow users to upload health data directly from their smartphones. • This will increase accessibility and enable users to monitor their health conditions anytime, anywhere, improving proactive healthcare management.

5. ACKNOWLEDGMENT

We extend our heartfelt gratitude to the researchers and professionals whose work laid the foundation for this study. Special thanks to [Nehru Arts and Science College] for providing the necessary resources and infrastructure for this research. We also acknowledge the invaluable guidance and support from our mentors and collaborators throughout this project. Finally, we are deeply grateful to the patients and medical professionals whose insights and data contributed to the development and validation of our findings..

REFERENCES

[1] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[2] Breiman, L. (2001). ”Random Forests.” Machine Learning, 45(1), 5-32.

[3] Chawla, N. V., Bowyer, K. W. (2002). ”SMOTE: Synthetic Minority Over-sampling Technique.” Journal of Artificial Intelligence Research, 16, 321-357.

[4] Chen, T., Guestrin, C. (2016). ”XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Confer- ence on Knowledge Discovery and Data Mining, 785-794.

[5] Dhar, V. (2018). ”Data Science and Prediction Models in Healthcare.” Journal of Data Science, 16(4), 35-48.

[6] Gonza´lez, R. C., Woods, R. E. (2002). Digital Image Processing. Prentice Hall.

[7] He, H., Wu, D. (2020). ”Cardiac Disease Diagnosis Using Machine Learning Algorithms.” Computational and Mathematical Methods in Medicine, 2020, 1-10.

[8] Khan, A. H., Chellappan, M. (2019). ”Machine Learning Techniques in Healthcare: A Survey and Classification.” Procedia Computer Science, 152, 117-124.

[9] Ng, A. Y. (2011). ”Machine Learning Yearning: Technical Strategy for AI Engineers.” Self-published.

[10] Peng X, Lu Z. Development and Validation of Upper Limb Lym- phedema in Patients After Breast Cancer Surgery Using a Practicable Machine Learning Model: A Retrospective Cohort Study. Int J Gen Med. 2024;17:3799-3812 https://doi.org/10.2147/IJGM.S478573.

[11] Rao, P. S., Thakur, M. (2019). ”Challenges and Trends in Machine Learning for Healthcare.” Journal of Healthcare Engineering, 2019, 1- 12.

[12] Sutton, R.T., Pincock, D., Baumgart, D.C. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digit. Med. 3, 17 (2020). https://doi.org/10.1038/s41746-020-0221-y

[13] Wang, Z., Gu, Y., Huang, L. et al. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data. Cardiovasc Diabetol 23, 351 (2024). https://doi.org/10.1186/s12933-024-02439-0